

For which countries did PPP hold? A multiple testing approach

Christoph Hanck

Received: 15 November 2007 / Revised: 15 March 2008 / Published online: 21 August 2008
© The Author(s) 2008. This article is published with open access at Springerlink.com

Abstract We employ a multiple testing technique to identify the countries for which purchasing power parity (PPP) held over the last century. The approach controls the multiplicity problem inherent in simultaneously testing for PPP on several time series, thereby avoiding spurious rejections. It has higher power than traditional multiple testing techniques by exploiting the dependence structure between the countries with a bootstrap approach. Our results show that, plausibly, thus controlling for multiplicity leads to a number of rejections of the null that is intermediate between that of traditional multiple testing techniques and that which results if one tests the null on each single time series at some level α .

Keywords Multiple testing · Bootstrap · PPP · Panel data

1 Introduction

Purchasing power parity (PPP) is among the most popular theories to explain the long run behavior of exchange rates. Not least because it is ready-made for empirical implementation, it has been investigated by a host of econometric techniques. So-called “stage-two” tests (Froot and Rogoff 1995) test the hypothesis that the real exchange rate follows a random walk. The alternative is that the real exchange rate is a stationary process, i.e., that PPP holds in the long run. Typically, researchers would obtain real exchange rate data over a certain time span for several countries and

Research supported by Ruhr Graduate School in Economics and DFG under Sonderforschungsbereich 475.

C. Hanck (✉)
Vogelpothsweg 78, 44221 Dortmund, Germany
e-mail: christoph.hanck@uni-dortmund.de

conduct appropriate unit root tests on each series (see, e.g., [Taylor 2002](#)). It is then argued that PPP holds for those countries for which the null is rejected.

Unfortunately, this simple and intuitive way of investigating the validity of PPP is problematic from a statistical point of view. Effectively, it ignores the issue of multiple testing. To illustrate the problem, consider the following artificial numerical example. Suppose one has exchange rate data on a panel of, say, $N = 20$ countries. Also assume for simplicity that the units are independent and that PPP does not hold for any of the units. When conducting tests on each unit at the $\alpha = 0.05$ level, one might casually expect the probability to erroneously find evidence in favor of PPP in at most one case to equal 5%, because $1/20 = 0.05$. However, the event of a rejection is a Bernoulli random variable with “success” probability 0.05. Hence, P_k , the probability of finding k rejections in N tests, is the probability mass function of a Binomial random variable,

$$P_k = \binom{N}{k} \alpha^k (1 - \alpha)^{N-k}.$$

Therefore, the probability of (at least) one erroneous rejection, also known as the Familywise Error Rate¹ ($FWER$), equals

$$P_{k \geq 1} = \sum_{j=1}^{20} \binom{20}{j} 0.05^j (1 - 0.05)^{20-j} = 0.6415.$$

Even if PPP does not hold for any of the countries in the panel, one will falsely find some evidence of it with a rather high probability. Of course, the problem only worsens if one adds more units to the panel.

This so-called “multiplicity” problem, while not widely recognized in econometrics ([Savin 1984](#)), has of course been realized long ago in the statistics literature (see [Lehmann and Romano 2005](#)). Several solutions to controlling the $FWER$ at some specified level α have been suggested. Among the most popular are the Bonferroni and the [Holm \(1979\)](#) procedure. These procedures have however been less successful in econometric applications because ensuring $FWER \leq \alpha$ typically comes at the price of reducing the ability to identify false hypotheses. That is, the procedures are conservative or have low “power.”² Hence, often quite reasonably, researchers have tended to ignore the issue of multiplicity.

Recently, panel econometric techniques have become popular to test for PPP. See, for instance, [Wu \(1996\)](#), [Papell and Theodoridis \(2001\)](#), [Papell \(2002\)](#) or [Murray and Papell \(2005\)](#). Typically, these panel unit root tests formulate the null of the entire panel being nonstationary. The alternative quite often is that of a stationary panel (see, for instance, [Harris and Tzavalis 1999](#); [Levin et al. 2002](#); [Breitung 2000](#)). These panel tests also have power against “mixed” panels, where only some fraction of the units is actually stationary (see [Taylor and Sarno 1998](#); [Karlsson and Löthgren 2000](#); [Boucher Breuer et al. 2001](#)). Hence, erroneous conclusions on the number of countries

¹ More generally, the j - $FWER$ is defined as $P_{k \geq j}$, the probability of j or more false rejections.

² For a discussion of “power” in a multiple testing framework see [Romano and Wolf \(2005, Sect. 2.2\)](#).

for which PPP holds remain possible. (Concluding from a rejection of a panel unit root test that *all* units are stationary is closely related to the erroneous inference that a rejection in an F test of the “significance of a regression” implies that *all* coefficients are nonzero.)

As a partial remedy, [Taylor and Sarno \(1998\)](#), [Im et al. \(2003\)](#), [Maddala and Wu \(1999\)](#) and [Choi \(2001\)](#) provide panel unit root tests having the more conservative alternative that some nonzero fraction of the panel is stationary. However, their approach neither allows to identify which nor how many of the countries in the panel have a stationary real exchange rate. [Boucher Breuer et al. \(2001\)](#) suggest a seemingly unrelated regressions ADF test to investigate how many series in the panel are stationary, without, however, taking the multiplicity problem into account.

Recently, there has been substantial research on improving the ability of multiple testing approaches to detect false hypotheses while still controlling the $FWER$. Notably, [Romano and Wolf \(2005\)](#) have put forward a bootstrap scheme that exploits the dependence structure of the statistics in order to improve the power of the multiple test. In the present paper, we propose an adaptation of the [Romano and Wolf \(2005\)](#) approach to identify those countries of a panel of real exchange rate data for which the PPP condition holds.

The plan of the paper is as follows. Section 2 offers a brief statement of the PPP condition and presents the general multiple testing approach of [Romano and Wolf \(2005\)](#). Section 3 discusses the bootstrap approach employed in this paper. The empirical results are in Sect. 4. Section 5 concludes.

2 The multiple testing approach

Our goal is to identify those countries of a panel for which the PPP relation held over the sample period. Let $p_{i,t}$ be the (log) price index in country i and period t , where $i = 1, \dots, N$ and $t = 1, \dots, T$, p_t^* the “foreign” (log) price index of the reference country in the panel and $s_{i,t}$ the (log) nominal exchange rate between the currencies of country i and the reference country. The real exchange rate is then given by

$$r_{i,t} = p_{i,t} - p_t^* - s_{i,t} \quad (i = 1, \dots, N).$$

Testing the strong PPP hypothesis is naturally formulated (see [Rogoff 1996](#)) as a unit root test on the real exchange rate. A vast number of unit root tests have been suggested in the literature (see [Phillips and Xiao 1998](#), for a survey), many of which have been applied to the PPP question. We will use the standard augmented [Dickey and Fuller \(1979\)](#) test (see also [Said and Dickey 1984](#)). We do so because it is still the most popular unit root test and, more importantly, the bootstrap versions of the test required for the multiple testing scheme have desirable properties ([Swensen 2003](#); [Chang and Park 2003](#)). Accordingly, we investigate PPP by testing the individual hypotheses

$$H_i : \varrho_i = 0 \quad \text{vs.} \quad H_i' : \varrho_i < 0 \quad (i = 1, \dots, N),$$

where

$$\Delta r_{i,t} = \mu_i + \varrho_i r_{i,t-1} + \sum_{j=1}^{J_i} \nu_j \Delta r_{i,t-j} + \epsilon_{i,t}. \quad (1)$$

The number of lagged differences J_i required to capture serial correlation in $r_{i,t}$, is allowed to vary across i . Our test statistic is given by $\hat{\tau}_i = \hat{\varrho}_i / s.e.(\hat{\varrho}_i)$, the t -statistic of ϱ_i in (1), where $\hat{\varrho}_i$ is the usual OLS estimator and $s.e.(\hat{\varrho}_i)$ the associated standard error.

We aim to determine those countries $i \in \{1, \dots, N\}$ for which $r_{i,t}$ is a stationary process. As argued in the Introduction, in order to provide reliable statistical inference in the sense of controlling the *FWER*, it is important to take into account the multiplicity inherent in testing in a panel setting. We now present the general multiple testing framework used here, making suitable adjustments to adapt the procedure to the PPP testing case.

First, relabel the test statistics from smallest to largest, such that $\hat{\tau}_{r_1} \leq \hat{\tau}_{r_2} \leq \dots \leq \hat{\tau}_{r_N}$. (The smaller a Dickey-Fuller test statistic, the stronger the evidence in favor of stationarity.) Form a joint rectangular confidence region for the vector $(\varrho_{r_1}, \dots, \varrho_{r_N})^\top$. The region is of the form

$$(-\infty, \hat{\varrho}_{r_1} + s.e.(\hat{\varrho}_{r_1}) \cdot d_1] \times \dots \times (-\infty, \hat{\varrho}_{r_N} + s.e.(\hat{\varrho}_{r_N}) \cdot d_1], \quad (2)$$

where one chooses d_1 so as to ensure a joint asymptotic coverage probability $1 - \alpha$.³ The bootstrap method to appropriately choose d_1 in the present problem will be discussed below. The decision rule is to reject a particular hypothesis H_{r_n} if the corresponding confidence interval satisfies $0 \notin (-\infty, \hat{\varrho}_{r_n} + s.e.(\hat{\varrho}_{r_n}) \cdot d_1]$. Romano and Wolf (2005) show that if the confidence region (2) has coverage probability $1 - \alpha$, then this method asymptotically controls the *FWER* at level α , $\lim_T FWER \leq \alpha$.⁴ Crucially, the method does not stop there. In order to improve the ability of the method to detect false hypotheses, one can construct further confidence regions after having rejected, say, the first N_1 hypotheses. In a second step, one forms a confidence region for the remaining $N - N_1$ coefficients $(\varrho_{r_{N_1+1}}, \dots, \varrho_{r_N})^\top$. This is again constructed to have nominal joint coverage probability $1 - \alpha$ and is of the form

$$(-\infty, \hat{\varrho}_{r_{N_1+1}} + s.e.(\hat{\varrho}_{r_{N_1+1}}) \cdot d_2] \times \dots \times (-\infty, \hat{\varrho}_{r_N} + s.e.(\hat{\varrho}_{r_N}) \cdot d_2],$$

potentially leading to the rejection of some further N_2 hypotheses. This step-down process can be repeated until no further hypotheses are rejected. Romano and Wolf (2005) show that the d_j should ideally be chosen as

³ As recommended by Romano and Wolf (2005) we use the studentized version of their method. For a discussion of the “basic” approach, see Sect. 3 of their paper.

⁴ They also provide ample Monte Carlo evidence of finite-sample effectiveness. In particular, the procedure is found to control the *FWER* also for finite T . Importantly, it is also shown to be quite powerful in that it successfully identifies a large number of false hypotheses.

$$d_j \equiv d_j(1 - \alpha, P) = \inf \left\{ x : \Pr_P \left[\max_{R_{j-1}+1 \leq n \leq N} \left(\frac{\hat{Q}_{r_n} - Q_{r_n}}{s.e.(\hat{Q}_{r_n})} \right) \leq x \right] \geq 1 - \alpha \right\},$$

where P denotes the true probability mechanism of the data, $R_{j-1} = \sum_{k=0}^{j-1} N_k$ and $R_0 = 0$. As such, $\Pr_P(A)$ is the probability of the event A computed under the true probability mechanism. In practice, however, P and hence d_j are unknown. Fortunately, [Romano and Wolf \(2005, Theorems 3.1 and 4.1\)](#) show that d_j can often be estimated consistently with the bootstrap without affecting asymptotic control of the *FWER*.

3 The bootstrap algorithm

We now outline the bootstrap approach to obtain an estimator \hat{d}_j employed in this paper.

1. Fit an autoregressive process to $\Delta r_{i,t}$ ($i = 1, \dots, N$; $t = 2, \dots, T$). It is natural to use the Yule–Walker procedure because it always yields an invertible representation ([Brockwell and Davis 1991](#), Sects. 8.1–2). Letting $\overline{\Delta r_i} := (T_i - 1)^{-1} \sum_{t=2}^{T_i} \Delta r_{i,t}$, compute the empirical autocovariances of $\Delta r_{i,t}$ up to order q ,

$$\hat{\gamma}_i(\ell) := \frac{1}{T_i - 1 - \ell} \sum_{t=2}^{T_i-\ell} (\Delta r_{i,t} - \overline{\Delta r_i})(\Delta r_{i,t+\ell} - \overline{\Delta r_i}),$$

where $i = 1, \dots, N$; $\ell = 0, \dots, q$.⁵ Defining

$$\hat{\Gamma}_i := \begin{pmatrix} \hat{\gamma}_i(0) & \cdots & \hat{\gamma}_i(q-1) \\ \vdots & \ddots & \vdots \\ \hat{\gamma}_i(q-1) & \cdots & \hat{\gamma}_i(0) \end{pmatrix}$$

and $\hat{\gamma}_i := (\hat{\gamma}_i(1), \dots, \hat{\gamma}_i(q))^\top$, obtain the AR coefficient vector as

$$(\hat{\phi}_{i,1}, \dots, \hat{\phi}_{i,q})^\top := \hat{\Gamma}_i^{-1} \hat{\gamma}_i \quad (i = 1, \dots, N).$$

2. The residuals are, as usual, given by

$$\hat{\epsilon}_{i,t} := \Delta r_{i,t} - \sum_{\ell=1}^q \hat{\phi}_{i,\ell} \Delta r_{i,t-\ell},$$

⁵ In practice, q can be chosen with a data-dependent criterion such as Akaike's.

for $i = 1, \dots, N$; $t = q + 2, \dots, T$. Following Swensen (2003), center $\hat{\epsilon}_{i,t}$ to obtain

$$\tilde{\epsilon}_{i,t} := \hat{\epsilon}_{i,t} - \frac{1}{T_i - q - 1} \sum_{g=q+2}^{T_i} \hat{\epsilon}_{i,g}$$

for $i = 1, \dots, N$; $t = q + 2, \dots, T$.

3. Resample (nonparametrically, with replacement) from $\tilde{\epsilon}_{i,t}$ to get $\epsilon_{i,t}^*$. To preserve the empirical cross-sectional dependence structure, jointly resample residual vectors

$$\tilde{\epsilon}_{\cdot,t} := (\tilde{\epsilon}_{1,t}, \dots, \tilde{\epsilon}_{N,t}). \quad (t = q + 2, \dots, T).$$

4. Recursively construct the bootstrap samples as⁶

$$\Delta r_{i,t}^* = \sum_{\ell=1}^q \hat{\phi}_{i,\ell} \Delta r_{i,t-\ell}^* + \epsilon_{i,t}^*$$

for $i = 1, \dots, N$, $t = q + 2, \dots, T$.

5. It is necessary to impose the null of a unit root when generating the artificial data in bootstrap unit root tests to achieve consistency (Basawa et al. 1991). Accordingly, impose the null of nonstationarity by integrating $\Delta r_{i,t}^*$ to obtain $r_{i,t}^*$.
6. For each bootstrap sample $r_b^* := \left((r_{b,1,1}^*, \dots, r_{b,1,T}^*)^\top, \dots, (r_{b,N,1}^*, \dots, r_{b,N,T}^*)^\top \right)$, compute the test statistics τ_{b,r_n}^* , and

$$\max_{b,j}^* := \max_{R_{j-1}+1 \leq n \leq N} (\tau_{b,r_n}^* - \hat{\tau}_{r_n}).$$

7. Repeat steps 3 to 6 many, say B , times.
8. Compute \hat{d}_j as the $1 - \alpha$ quantile of the B values $\max_{1,j}^*, \dots, \max_{B,j}^*$.

Chang and Park (2003) and Swensen (2003) show that the above sieve bootstrap scheme yields asymptotically valid bootstrap ADF tests in the sense that using the α quantile of the bootstrap distribution of the τ_{b,r_n}^* as critical value asymptotically gives a test with size α . By a continuous mapping theorem argument, we expect the bootstrap to also consistently estimate the distribution of the $\max_{b,j}^*$ and hence \hat{d}_j .⁷

4 Results

We now present the empirical results of an application of the modified Romano and Wolf (2005) methodology to the PPP condition. We revisit the dataset used by Taylor

⁶ We run the recursion for 30 initial observations before using the $\Delta r_{i,t}^*$ to mitigate the effect of initial conditions.

⁷ For a similar argument, see White (2000, p. 1104)

Table 1 Empirical results

	Country	$\hat{\tau}_i$	p -value	Holm criterion
	Mexico	-4.334	< 0.001	0.0026
	Finland	-4.136	0.001	0.0028
	Argentina	-3.632	0.006	0.0029
	Italy	-3.344	0.015	0.0031
	Norway	-3.285	0.018	0.0033
	Sweden	-3.202	0.022	0.0036
	UK	-2.996	0.038	0.0038
	Belgium	-2.980	0.040	0.0042
	Germany	-2.957	0.042	0.0046
	France	-2.929	0.045	0.0050
	Brazil	-2.561	0.104	0.0056
	Australia	-2.544	0.108	0.0063
	Netherlands	-2.498	0.119	0.0071
$\hat{\tau}_i$ are unit-root t -test statistics	Portugal	-2.391	0.147	0.0083
on the real exchange rate,	Canada	-2.202	0.207	0.0100
p -value are the corresponding	Spain	-2.118	0.238	0.0125
marginal significance levels and	Denmark	-2.058	0.262	0.0167
Holm criterion is the value that,	Switzerland	-1.349	0.604	0.0250
once exceeded by the adjacent	Japan	-1.323	0.617	0.0500
p -value, implies no rejection for				
all units having larger p -values				

(2002), which includes annual data for the nominal exchange rate, CPI and the GDP deflator. This dataset is particularly useful for our purposes because it covers a long period, ranging from 1892 through to 1996. The countries contained in our panel are given in Table 1. We use the United States as the reference country and report results using CPI price series.⁸ See Taylor (2002) for further details on data sources and definitions.

Using standard ADF unit root tests, we find rejections for 10 out of 19 countries at the MacKinnon (1996)-type simulated finite sample 5% critical value -2.889 . See the first column of Table 1. (The entries are sorted for later use.) The number of lagged differences J_i in (1) is chosen with the data-dependent criterion of Ng and Perron (2001). The findings of Taylor (2002) are very similar.⁹ Evidence in favor of PPP is therefore at best mixed. Taylor (2002) then argues that it may be possible to find more rejections in favor of PPP by employing more powerful techniques. Our goal, on the other hand, is to investigate whether some of the rejections are spurious in the sense that they would not occur when taking into account the multiplicity of the testing problem.

As a preliminary step, we report results for more classical techniques to control the FWER, namely the Bonferroni and the Holm (1979) procedures. Recall that the

⁸ To check whether our procedure is robust to re-normalization, we also perform the analysis using the United Kingdom as reference country. Detailed results, which are very similar to those to be reported below, are available upon request.

⁹ The small differences can be explained by different interpolation schemes for missing wartime data, other lag selection criteria as well as the fact that we balance our panel.

Table 2 Panel unit root test results

		Bootstrap c.v.
* Significance at the 5% level. P_m is a test from Choi (2006) , PS_{Fisher} from Phillips and Sul (2003) , t_{rob} from Breitung and Das (2005) , DHT from Demetrescu et al. (2006) and $CIPS$ is by Pesaran (2007) . K_{OLS} and K_{GLS} are two of Chang's (2004) tests	P_m	23.888*
	PS_{Fisher}	969.802*
	t_{rob}	-3.618*
	DHT	-7.882*
	$CIPS$	-3.172*
	K_{OLS}	106.833* 74.440
	K_{GLS}	195.448* 130.096

former rejects H_i if the p -value \hat{p}_i corresponding to the test statistic \hat{t}_i satisfies $\hat{p}_i \leq \alpha/N$. The [Holm \(1979\)](#) procedure first sorts the p -values from smallest to largest, $\hat{p}_{r_1} \leq \dots \leq \hat{p}_{r_N}$. Relabel the hypotheses accordingly as H_{r_n} . Then, reject H_{r_n} at level α if $\hat{p}_{r_j} \leq \alpha/(N - j + 1)$ for all $r_j = 1, \dots, r_n$. The cutoff value for the first hypothesis is identical for both methods, but unlike the Bonferroni method, the [Holm \(1979\)](#) procedure uses gradually less challenging criteria for H_{r_2}, \dots, H_{r_N} . Nevertheless, it often has low power because it also fails to exploit the dependence structure between the statistics.

The limit distribution of the ADF test statistics is a functional of Brownian motions that cannot be evaluated analytically to obtain p -values. We therefore rely on response surface regressions suggested by [MacKinnon \(1994, 1996\)](#) to obtain numerical distribution functions of the test statistics. We report results in columns 2 and 3 of Table 1.

As expected, the number of rejections is now much lower. After controlling for multiplicity, we only observe rejections for Mexico and Finland for either method. These results indeed suggest that the Bonferroni and Holm procedures are conservative.

As discussed in the Introduction, panel unit root tests are a competing means to analyze whether PPP holds in a panel dataset. For comparison, we therefore report results for some popular tests. It is known since at least [O'Connell \(1998\)](#) that panels of real exchange rates are typically cross-sectionally dependent.¹⁰ Thus, we focus on so called second generation tests, which do not require the assumption of cross-sectional independence. Results are reported in Table 2. Clearly, all tests reject the panel unit root null, perhaps unsurprisingly in view of the substantial number of rejections when using time series unit root tests (see Table 1). These results indicate that PPP holds for a nonzero fraction of the countries in our dataset. As discussed in the Introduction, it would however be misleading to conclude that PPP holds for all countries. Also, the results are not informative about for which countries PPP can be said to hold.

We therefore now turn to the results of the [Romano and Wolf \(2005\)](#) approach. The algorithm presented in Sect. 2 yields $\hat{d}_1 = 4.050$, leading to a rejection for Mexico and Finland. In the second round, we obtain $\hat{d}_2 = 3.429$, implying evidence in favor of PPP for Argentina. Next, we find $\hat{d}_3 = 3.252$ such that we reject for Italy and Norway. Finally, $\hat{d}_4 = 3.075$ means that we also reject the null in the case of Sweden.

¹⁰ Indeed, a CD test for cross section dependence as discussed in [Pesaran et al. \(2008\)](#) emphatically rejects the null of cross section independence. The p -value is indistinguishable from zero.

Observe that the number of rejections is intermediate between the results for the Holm and Bonferroni methods and that of the individual country results. In view of the above discussion, we find that this result is rather plausible. Furthermore, the ability of the [Romano and Wolf \(2005\)](#) method to detect several false hypotheses in a stepwise fashion proved instrumental in improving upon the more traditional multiple testing methods.

5 Conclusion

We have used recent advances in the multiple testing literature to make an attempt to identify those countries for which PPP held over the last century. The approach controls the multiplicity problem inherent in simultaneously testing for PPP on several time series, thereby avoiding spurious rejections. It has higher power than traditional multiple testing techniques by exploiting the dependence structure between the countries with a bootstrap approach. We use a sieve bootstrap approach to account for nonstationarity under the null hypothesis. On the other hand, our empirical results show that, plausibly, controlling for multiplicity leads to fewer rejections of the null of no PPP than if one tests the null on each single time series at some level α . Specifically, we find rejections of the null of no PPP for Mexico, Finland, Argentina, Italy, Norway and Sweden.

Obviously, the present framework is fairly general and could be applied to other macroeconomic questions such as savings-investment correlation or spot and forward exchange rates ([Mark et al. 2005](#)) that have hitherto been dealt with using panel techniques. Another fruitful task for future research would be to apply the present framework using more powerful unit root tests such as those by [Elliott et al. \(1996\)](#), in order to improve the ability of the procedure to identify false hypotheses. Also, it is possible to accommodate problems that imply testing for cointegration.

Acknowledgments I am indebted to two anonymous referees and an associate editor whose remarks substantially improved the paper as well as to Alan Taylor for providing the dataset used in this paper. I am grateful to conference participants at Journée d'économétrie, Paris, DIW Macroeconometric Workshop, Berlin, Italian Congress of Econometrics, Rimini, and Verein für Socialpolitik, München, as well as seminar participants at Universities of Dortmund, Münster, and Maastricht, for helpful discussion.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Basawa IV, Malik AK, McCormick WP, Reeves JH, Taylor RL (1991) Bootstrapping unstable first-order autoregressive processes. *Ann Stat* 19:1098–1101
- Boucher Breuer J, McNown R, Wallace MS (2001) Misleading inferences from panel unit root tests with an illustration from purchasing power parity. *Rev Int Econ* 9:482–493
- Breitung J (2000) The local power of some unit root tests for panel data. In: Baltagi BH (ed) *Nonstationary panels, panel cointegration, and dynamic panels*. Elsevier, Amsterdam, pp 161–177

- Breitung J, Das S (2005) Panel unit root tests under cross sectional dependence. *Stat Neerl* 59:414–433
- Brockwell PJ, Davis RA (1991) Time series: theory and methods, 2nd edn. Springer, New York
- Chang Y (2004) Bootstrap unit root tests in panels with cross-sectional dependency. *J Econom* 120:263–293
- Chang Y, Park JY (2003) A sieve bootstrap for the test of a unit root. *J Time Ser Anal* 24:379–400
- Choi I (2001) Unit root tests for panel data. *J Int Money Finance* 20:249–272
- Choi I (2006) Combination unit root tests for cross-sectionally correlated panels. In: Corbae D, Durlauf SN, Hansen BE (eds) *Econometric theory and practice: frontiers of analysis and applied research: essays in honor of Peter C. B. Phillips*, chap 11. Cambridge University Press, Cambridge, pp 311–333
- Demetrescu M, Hassler U, Tarcolea AI (2006) Combining significance of correlated statistics with application to panel data. *Oxf Bull Econ Stat* 68:647–663
- Dickey DA, Fuller WA (1979) Distribution of the estimators for autoregressive time series with a unit root. *J Am Statist Assoc* 74:427–431
- Elliott G, Rothenberg TJ, Stock JH (1996) Efficient tests for an autoregressive unit root. *Econometrica* 64:813–836
- Froot KA, Rogoff K (1995) Perspectives on ppp and long-run real exchange rates. In: Grossman G, Rogoff K (eds) *The handbook of international economics*, vol 3, chap 32. Elsevier, Amsterdam, pp 1647–1688
- Harris RD, Tzavalis E (1999) Inference for unit roots in dynamic panels where the time dimension is fixed. *J Econom* 91:201–226
- Holm S (1979) A simple sequentially rejective multiple test procedure. *Scand J Stat* 6:65–70
- Im KS, Pesaran MH, Shin Y (2003) Testing for unit roots in heterogeneous panels. *J Econom* 115:53–74
- Karlsson S, Löthgren M (2000) On the power and interpretation of panel unit root tests. *Econ Lett* 66:249–255
- Lehmann EL, Romano JP (2005) Testing statistical hypotheses, 3rd edn. Springer, New York
- Levin A, Lin CF, Chu CSJ (2002) Unit root tests in panel data: asymptotic and finite-sample properties. *J Econom* 108:1–24
- MacKinnon JG (1994) Approximate asymptotic distribution functions for unit-root and cointegration tests. *J Bus Econ Stat* 12:167–176
- MacKinnon JG (1996) Numerical distribution functions for unit root and cointegration tests. *J Appl Econom* 11:601–618
- Maddala G, Wu S (1999) A comparative study of unit root tests with panel data and a new simple test. *Oxf Bull Econ Stat* 61:631–652
- Mark NC, Ogaki M, Sul D (2005) Dynamic seemingly unrelated cointegrating regression. *Rev Econ Stud* 72:797–820
- Murray CJ, Papell DH (2005) Do panels help solve the purchasing power parity puzzle. *J Bus Econ Stat* 23:410–415
- Ng S, Perron P (2001) Lag length selection and the construction of unit root tests with good size and power. *Econometrica* 69:1519–1554
- O’Connell PG (1998) The overvaluation of purchasing power parity. *J Int Econ* 44:1–19
- Papell DH (2002) The great appreciation, the great depreciation, and the purchasing power parity hypothesis. *J Int Econ* 57:51–82
- Papell DH, Theodoridis H (2001) The choice of numeraire currency in panel tests of purchasing power parity. *J Money Credit Bank* 33:790–803
- Pesaran MH (2007) A simple panel unit root test in the presence of cross section dependence. *J Appl Econom* 22:265–312
- Pesaran MH, Ullah A, Yamagata T (2008) A bias-adjusted lm test of error cross section independence. *Econom J* 11:105–127
- Phillips PCB, Sul D (2003) Dynamic panel estimation and homogeneity testing under cross section dependence. *Econom J* 6:217–259
- Phillips PCB, Xiao Z (1998) A primer on unit root testing. *J Econ Surveys* 12:423–469
- Rogoff K (1996) The purchasing power parity puzzle. *J Econ Lit* 34:647–668
- Romano JP, Wolf M (2005) Stepwise multiple testing as formalized data snooping. *Econometrica* 73:1237–1282
- Said SE, Dickey DA (1984) Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* 71:599–607
- Savin NE (1984) Multiple hypothesis testing. In: Griliches Z, Intriligator M (eds) *Handbook of econometrics*, vol 2, chap 14. North-Holland, Amsterdam, pp 827–879
- Swensen AR (2003) Bootstrapping unit root tests for integrated processes. *J Time Ser Anal* 24:99–126
- Taylor AM (2002) A century of purchasing-power parity. *Rev Econ Stat* 84:139–150

- Taylor MP, Sarno L (1998) The behavior of real exchange rates during the post-bretton woods period. *J Int Econ* 46:281–312
- White H (2000) A reality check for data snooping. *Econometrica* 68:1097–1126
- Wu Y (1996) Are real exchange rates stationary? evidence from a panel data test. *J Money Credit Bank* 28:54–63